Response to Anonymous Referee #1

First of all, we would like to thank referee #1 for his valuable comments on our manuscript. We used italics to mark our answers to his comments.

General comments:

The authors in section 2.3 mentioned that there was no need for them to calibrate the model since it was an impact study. I strongly disagree because they still need to assess model performance with observed data and that can be done with site specific calibration. This will greatly influence the impact assessment results.

Answer: The model was parameterized using GIS databases, literature values, default parameters, and a dam management scenario for the reservoirs. A detailed overview of the model parameterization is presented in Wagner et al. (2011; 2012). The validation of the model with measured stream flow data yielded Nash-Sutcliff efficiencies better than 0.67 (gauge G1: 0.68 to 0.69, gauge G4: 0.67) and a percentage bias between 4% and 24% (G1: +4% to +5%, G4: +24%). A more detailed validation that includes hydrographs is also published in Wagner et al. (2011; 2012).

The indices of agreement provide evidence for a suitable model parameterization. While there might be a potential to improve the agreement of the model results with the measurements by a specific calibration, there is no evidence, that these optimized parameters are in fact the correct parameters, since this parameter optimization and the interactions of different model parameters might as well lead to better results for the wrong reason. Thus we abstained from a site specific statistical calibration and used a parameterization based on the available data instead. However, some model parameters taken from the literature are in fact regionally adapted, e.g. soil available water capacity was taken from Immerzeel et al. (2008), who calibrated this value in their larger scale study on the same region.

We agree with the reasoning of Kirchner (2006) that tuning the model to one specific condition may introduce a bias when the model is applied to other conditions. I.e., if both models were calibrated for the respective land use map, it is possibly or even likely that the changes induced by the different land use inputs will be masked by a different type of calibration.

We will clarify this reasoning in the revised manuscript.

Reference: [Kirchner, J. W.: Getting the right answers for the right reasons: Linking measurements, analyses, and models to advance the science of hydrology. Water Resour. Res., 42, W03S04, 2006.]

On section 3.2 the authors, should in addition to correlations, carry out trend analyses on rainfall and simulated flows. The reason for this is to ensure that there are no false signals of land use change, which may have been attributed by rainfall changes. These results should be presented. Also the significance of change should be quantified using for instance parametric or non-parametric tests.

The model runs for both land use classifications were performed using the same weather input data. The results were compared using average values of water balance components for the 20-year model runs. Thus, effects of inter-annual weather changes (even if they had a

significant trend) do not affect our analysis, which is focused on average values and relative changes between the two model runs. Hence, we did not apply these additional analyses.

Otherwise, it is an interesting study with good assessment on classification of land uses.

Thank you.